Improved Method for Neural Spike Alignment: The Centroid Filter

B.W. Metcalfe[a,*], C.T. Clarke[a], N. de N. Donaldson[b], J.T Taylor[a]

[a]*Department of Electronic and Electrical Engineering, University of Bath, Bath UK*
[b]*Department of Medical Physics and Biomedical Engineering, University College London, London UK*

Highlights

- A key problem in neural spike sorting is highlighted.
- A new method to measure the position of action potentials in real-time is described.
- This method is validated using detailed simulations.
- Major improvements are demonstrated compared to standard methods.

Abstract

Background: This paper describes a new method to calculate the positions of *action potentials* (APs) in extracellular neural recordings in real-time. Extracellular recordings made using various types of electrode structures are a key tool in experimental electrophysiology and involve the identification of signals from individual neurons, a process known by the generic name *spike sorting*. An essential feature of the process is the alignment of individual APs ('spikes') in time. Given the recordings are invariably made in noisy environments there is a tendency to timing jitter leading to alignment errors. This is especially true if single point methods of AP location are employed.

New Method: The problems of single point AP location in the presence of noise are significantly reduced by the use of the *centroid* as a measure of position, which is based on a time average of all the points in the AP. The method is described and analysed in detail and it is shown that a very economical hardware realisation is possible.

Results: The new methods are investigated by simulation using deterministic models of nerve signals (i.e. APs with added noise; both correlated and uncorrelated noise models are considered) as well as results measured from a hardware implementation. A power efficient realisation of the centroid filter is demonstrated using a *complex programmable logic device* (CPLD).

Comparison with Existing Method: Detailed simulations demonstrate the superiority of the centroid method compared to several standard, single point metrics.

Conclusions: A new method has been suggested that significantly improves one of the key issues in neural spike sorting. The technique has the potential to influence significantly the design of electrophysiological recording systems in the future.

Keywords: action potentials, spike sorting, electrophysiology, centroid filter

1. INTRODUCTION

Extracellular recording of physiological neural activity is the dominant experimental technique in electrophysiology providing valuable information about both the central and peripheral nervous systems [1]. Extracellular electrodes such as hooks or suction electrodes are considered to be both reliable and minimally invasive recording approaches. These types of electrode spontaneously record the electrical activity from an unknown number of excited may serve different functions such as movement or memory. The study of these high-order functions often requires simultaneous recording from large areas of interconnected neurons using, for example, multi-channel electrode arrays [2]. There is a special interest in reconstructing the waveform of individual neurons from these multi-channel recordings so that the firing rates, or spike trains, can be extracted and compared. This procedure is referred to as neural *spike sorting* which is the labelling and classification of individual *action potentials* (AP) based on both shape and amplitude. It is generally considered that the shape and amplitude of APs recorded from a single neuron are time invariant and are a function of the axon diameter and the distance from the neuron to the recording electrode.

Historically the spike sorting process was performed visually by a researcher who, after first determining the number of different classes, would classify each AP based on shape and reconstruct individual spike trains for each neuron. This time-consuming process has since been replaced by more reliable computer based, automated methods that make use of statistical analysis and signal processing [3], [4], [5]. There are many different approaches to automated spike sorting but the majority can be broken down into *seven* main steps [6]:

1. **Filtering** – In the first stage the raw data recorded from a single extracellular electrode is band-pass filtered to remove the frequency components of interference and noise that are out of band. The frequency range of interest is typically 10 – 10,000 Hz.
2. **Spike Detection** – After filtering individual APs need to be identified. Nearly all methods detect APs with two main steps; the pre-emphasis of the signal and the use of an amplitude threshold. Pre-emphasis may be performed using methods such as the *non-linear energy operator* that attempts to locate rapid changes within the signal [7]. The amplitude threshold is generally determined automatically using an un-supervised training process [8].

3. **Alignment** – Once APs have been identified within the recording a window of data is extracted and the waveforms are captured. At this point each AP is aligned based on a measure such as the maximum value or the point of maximal slope [9].
4. **Feature Extraction** – Often very simple features such as the maximum amplitude are used to describe the morphology of the AP [9]. More complex methods employ statistical concepts such as *principal component analysis* (PCA) to extract unique features from the captured AP [10].
5. **Dimensionality Reduction** – The resulting features have many dimensions that make identification computationally expensive; one method to reduce the dimensionality of the features is simply to subsample at regular intervals. (there are more advanced statistical methods such a Hartigan's dip test [11], [12]). In PCA it is common to reduce the number of dimensions considered for clustering by simply truncating to *N* dimensions, where *N* is typically 2 or 3 [13].
6. **Clustering** – The reduced data set can now be automatically examined for clusters that correspond to different classes. Clustering (especially un-supervised) clustering is often the most difficult part of the sorting process. The *de-facto* benchmark is *k*-means clustering but this relies directly on human intervention and so is not suitable for un-supervised learning. One algorithm that is suitable for un-supervised clustering is called *valley seeking* [14], however the algorithm is not real time and has serious drawbacks in terms of complexity. When dealing with clusters that are irregular in shape, improvements may be obtained with *superparamagnetic* clustering [1].
7. **Classification** – As new APs are observed in the recorded data the clusters are updated and can be used to classify new APs. From this a visual record called a *spike train* can be produced that shows the firing patterns of each class of AP.

There are two main sources of spurious signals that pose significant challenges in extracellular recordings: deterministic common mode interference from sources such as excited muscles and random signals such as thermal noise. When noise and interference are present, similar APs originating from different neurons may appear to be the same or, alternatively, APs from the same neuron may appear different. One of the critical constraints of the spike sorting process is the alignment of individual APs before feature extraction is applied. Any temporal misalignment of APs at this time can have a detrimental effect on spike classification. The alignment process is often

performed at higher sampling rates in order to minimise sampling jitter and the most common method is to align each AP to the point of *maximum slope* [9]. Other methods consider the point of maximum amplitude or the point (or points) at which the amplitude has been reduced by 3 dB. These methods typically employ only one or two samples from the entire AP and so are highly sensitive to the broadband noise that is typical of biological recordings [15]. Alignment to a metric that is derived from the whole AP rather than from a single point will in general be less susceptible to the effects of both background noise and interference and it is this approach that is considered in this paper.

Section 2 of the paper begins with a brief overview of the existing methods for spike alignment and describes some of the basic limitations of these methods, in particular the effect of noise on the alignment process. A novel real time approach for spike alignment based on a *centroid* filter that provides an alternative to the traditional spike alignment methods and substantially improves the resilience to noise and sampling jitter is then described (Section 3). Validation of the new methods is achieved by simulation using deterministic models of nerve signals (i.e. APs with added noise; both correlated and uncorrelated noise models are considered) as well as results measured from a hardware implementation. Comparisons are made between the methods. A power efficient realisation of the centroid filter is described that operates in real time on a single Altera Max V *complex programmable logic device* (CPLD). The use of CPLDs for signal processing tasks is well documented and they are particularly well suited to the rapid analysis and assessment of novel techniques. A CPLD is used here to evaluate the power and resource requirements for the proposed designs that are key metrics for implantable technologies [16] (Section 4).

2. CENTROID FILTERING TECHNIQUE

In this section we describe a more robust and effective method that performs temporal alignment of APs based on an average of all of the samples within the AP. If the AP is considered to be a topological shape (i.e. a closed non-self-intersecting polygon) then the centre of the AP in both time and amplitude can be considered to be the *centroid* [17][18]. This concept is illustrated in Figure 1 where a waveform (solid line) has been partitioned into the shaded area that can be treated as a plane polygon with centroid located at C_x, C_y, this is the centre of mass of the hypothetical polygon.

Because the waveform is considered as a plane figure, the time axis is labelled *x* and the voltage axis *y*. We regard it as a polygon because the continuous time waveform has been sampled. Alignment to the centroid has the potential to be substantially more robust to noise and interference. Consider now the hypothetical time domain AP shown in the solid trace of Figure 2. The AP can be partitioned into a polygon bounded by its width at the points ω_1 and ω_2 and by the interception with the line $y = 0$, forming a plane polygon with unit density per unit area. In practice the interception with the line $y = 0$ may be found by half-wave rectification of the waveform. The centroid of the AP calculated along the *x* axis can therefore be found by using standard methods, i.e. taking the first moment of area about the axis and dividing by the total area [19]:

$$C_x \stackrel{\text{def}}{=} \frac{1}{A}\int_{\omega_1}^{\omega_2} x f(x)dx \tag{1}$$

where the centroid calculated along the *x* axis is C_x, A is the area under the AP, $(\omega_2 - \omega_1)$ is the width of the AP and $f(x)$ represents the AP function itself. Consider next the *convolution integral* of *f(x)* with another function *h(x):*

$$y(x) = (h * f)(x) = \int_{-\infty}^{\infty} h(x-\lambda)\, f(\lambda)d\lambda \tag{2}$$

Note that if this integral is evaluated at the origin ($x = 0$), after a change of variables and a suitable choice of limits, it reduces to:

$$y(0) = \int_{-\infty}^{\infty} h(-\lambda)\, f(\lambda)d\lambda \tag{3}$$

Furthermore if *h(x)* is chosen to be *kx*, where *k* is a negative constant, then Equation (3) reduces to Equation (1) (with the exception of the 1/*A* scaling term that is not required if the integral is evaluated at the origin). To illustrate this process consider the test pulse $f(x)$ shown in Figure 3 (dashed 'top hat' function of unit area). Evaluating the integral in Equation (1) where the limits of the integration are the dimensions of *f(x)* along the *x* axis ($\omega_1 = -0.5$, $\omega_2 = 0.5$) the centroid can be shown to be at the origin ($C_x = 0$).

For the convolution method, as noted above, we choose *h(x)* to be a linear function of *x* with a negative gradient passing through the origin. The width of the function *h* (i.e. the points at which $h = \pm 1$) is chosen to be greater than the width of *f*, for reasons that

will become apparent shortly. In this example the constant *N* was chosen to be 1.5, as indicated in Figure 3:

$$h(x) = -\frac{2}{N}x \tag{4}$$

For values of *x* where the two functions do not overlap the product of $(h * f)(x)$ is zero. Where overlap occurs, since both functions are discontinuous, a piecewise approach is taken. There are three phases to be considered, depending on the extent of the overlap of the two functions, as $h(x - \lambda)$ approaches $f(x)$ from the left. The first phase occurs as *h* approaches ω_1, which is -0.5 in this case and terminates when the leading edge of *h* is coincident with ω_2 (0.5):

$$\begin{aligned}(h * f)_1(x) &= \frac{-2}{N}\int_{\omega_1}^{x+0.5N}(x-\lambda)\,d\lambda \\ &= \frac{0.25N^2 - (x-\omega_1)^2}{N}\end{aligned} \tag{5}$$

This function is a parabola symmetrical about the vertical axis $x = \omega_1$ (-0.5 in this case) and displaced vertically by 0.25 *N*, as indicated in normalised form in the lower plot in Figure 3. Similarly the third phase of the process ends when the trailing edge of *h* makes its last contact with ω_2 (i.e. 0.5) and is given by the following integral:

$$\begin{aligned}(h * f)_3(x) &= \frac{-2}{N}\int_{x-0.5N}^{\omega_2}(x-\lambda)\,d\lambda \\ &= \frac{-0.25N^2 + (-x+\omega_2)^2}{N}\end{aligned} \tag{6}$$

Between phases 1 and 3, during phase 2, *h* completely encloses *f* since *N* was chosen to be greater than $|\omega_2 - \omega_1|$. In this phase the convolution integral is a linear function:

$$\begin{aligned}(h * f)_2(x) &= \frac{-2}{N}\int_{\omega_1}^{\omega_2}(x-\lambda)\,d\lambda \\ &= \frac{(\omega_2-\omega_1)}{N}[-2x + (\omega_2+\omega_1)]\end{aligned} \tag{7}$$

passing through the point $x = (\omega_2 + \omega_1)/2$, which is the centroid of $f(x)$. So in the example shown in Figure 3, where $f(x)$ is symmetric about the origin in *x* ($\omega_1 = -\omega_2$), the convolution integral is simply a straight line passing through the origin with gradient -2/*N*. In addition, from Equations (5)-(7) it is easy to show that the parabolas

of phases 1 and 3 intersect with the linear phase 2 at the following points on the x axis:

$x_{12} = -0.5N - \omega_1$ (intersection of phases 1 & 2)

$x_{23} = 0.5N - \omega_2$ (intersection of phases 2 & 3)

For the example in Figure 3 these values of x are -0.25 and 0.25, respectively, as can be seen clearly in the Figure. These equations also show the significance of the earlier observation that the function h must be wider than f. In the limit where the widths of the functions are equal, the intersection points x_{12} and x_{23} are both zero and the linear phase 2 disappears. In this case it is *not* possible to use the method to determine the centroid of $f(x)$. Of course if the width of $h(x)$ is increased to be much wider than $f(x)$ then it will find the centroid of the combination of $f(x)$ and any noise either side of $f(x)$. Since $h(x)$ can be realised as an FIR filter of length N, given that the relative length condition is satisfied, a single filtering operation on $f(x)$, plus zero crossing detection, is sufficient to calculate its centroid. Returning to the example shown in Figure 3, the result of the filter operation on a single realistic AP (solid line) is shown, the output from the filter crosses the origin at a point coincident with the centroid of the AP (dashed line). Note that the output from the filter goes negative after zero crossing, this corresponds to third phase of the process as described in Equation (6).

The general form of the output expression $y[n]$ for an FIR filter of length N in the time domain with inputs $x[n]$ and coefficients b_i is as follows:

$$y[n] = b_0.x[n] + b_1.x[n-1] + \cdots + b_N.x[n-N] = \sum_{i=0}^{N} b_i.x[n-i] \qquad (8)$$

To realise Equation (4) as an FIR filter we set:

$$b_i = -\frac{2}{N}i + 1 = mi + 1 \qquad (9)$$

i.e. $b_0 = 1$ and $b_N = -1$ and $m = -2/N$ and where the constant (unity) term is required to make the filter realisable. However, substitution of Equation (9) into Equations (5) - (7) shows that the modified form of h introduces an output delay proportional to $N/2$ that offsets the position of the zero crossing. This can be compensated for by using a shift register of the same length to delay the input signal before processing.

3. IMPLEMENTATION AND OPTIMISATION

The centroid filter discussed in Section 2 can now be optimised for implementation in VLSI architectures where power consumption is a prime concern. In a direct form implementation, a length *N* FIR filter requires *N* multipliers and *N* adders and generally speaking multiplication will dominate the power budget. In this section we describe a method that uses the linear form of Equation (9) to advantage in order to avoid the use of multipliers as far as possible. We begin by expanding Equation (8) and hence calculate the next output (Equation (11)):

$$y[n+1] = b_0.x[n+1] + \cdots + b_N.x[n-N+1] \ = \sum_{i=0}^{N} b_i.x[n+1-i] \qquad (10)$$

Recalling that $b_0 = 1$ and $b_N = -1$, then, from Equation (9):

$$b_i = mi + 1 \qquad (11)$$

and the following recurrence relation between the coefficients can be derived:

$$b_i - b_{i-1} = m \qquad (12)$$

Using this expression, Equation (10) for the next output value can be rearranged as follows:

$$y[n+1] = y[n] + \left\{ m \sum_{i=0}^{N-1} x[n-i] \right\} + x[n+1] \ - x[n-N] \qquad (13)$$

The computation of $y[n + 1]$ in this way requires $N + 3$ additions and one multiplication. Furthermore, if *m* is a power of 2 the multiplication can be replaced by a left shift. Also, the computation of the summation term in Equation (13) can be accomplished by using a *rolling sum* expressed as follows where $\omega[n]$ is the rolling sum:

$$\omega[n] = m \sum_{i=0}^{N-1} x[n-i] \qquad (14)$$

$\omega[n]$ can be expanded as follows:

$$\omega[n+1] = \ \omega[n] + m\{x[n+1] - x[n-N-1]\} \qquad (15)$$

Substitution of Equation (15) back into Equation (13) results in the following

optimised expression for $y[n + 1]$:

$$y[n+1] = y[n] + \omega[n] + m\{x[n+1] - x[n-N-1]\} + x[n+1] - x[n-N] \quad (16)$$

The computational effort has been reduced drastically compared to a direct implementation of an FIR filter and an implementation of Equation (16) requires *one* multiplier and *five* adders regardless of the length of the filter (the earlier remark about *m* being chosen as a power of 2 still applies). The shift register used to store the previous samples is the only component that is dependent on the order of the filter.

4. Modelling and Simulated Results

This section provides validation of the new methods using simulations based on deterministic models of the AP and noise functions.

A. *Single Fiber Action Potential Model*

The TMAP model used for generating SFAPs is described in Equation (17). It is a first-order analytical model of the extracellular AP and is based on a damped sinusoidal function as proposed in [13], [20].

$$f(x) = A \sin\left(\frac{x}{\tau_1}\right) \mathrm{e}^{-\frac{x}{\tau_2}} \quad (17)$$

The parameters A, τ_1 and τ_2 determine the amplitude, rising edge rate and duration of each SFAP, respectively. The parameters used are listed in Table I and were given in the literature as biologically plausible parameters for mammalian axons based on recordings made in *frog* [20]. The relationship between the axonal diameter and the conduction velocity can be loosely expressed by:

$$v = d(0.06T + 0.6) \quad (18)$$

TABLE I

MODEL PARAMETERS THAT DEFINE THE SFAP FUNCTIONS COMPUTED USING EQUATION (17).

Axon Diameter (μm)	Constant A	τ_1(ms)	τ_2(ms)
5	2.42	0.175	0.25
7	2.65	0.120	0.15
9	2.73	0.093	0.11
11	2.73	0.080	0.096
13	2.79	0.078	0.092
15	2.80	0.076	0.089
19	2.89	0.072	0.084

where *v* is the conduction velocity, *d* is the diameter of the fibre and *T* is the absolute temperature [20]. Figure 4 shows seven SFAPs generated using this model, where the parameters have been extracted from axons of different diameters (and thus conduction velocities).

B. Noise models

In each simulation random noise was added to noiseless APs to produce a signal with a specific, controllable signal-to-noise ratio (SNR). The following standard formula was employed to calculate SNR:

$$SNR(db) = 10log\left(\frac{P_{signal}}{P_{noise}}\right) = 10log\left(\frac{mean_{signal}^2+\sigma_{signal}^2}{mean_{noise}^2+\sigma_{noise}^2}\right) \quad (19)$$

Where the SNR was measured over the entire simulated recording, which lasted $100\,ms$. Both *additive white Gaussian noise* (AWGN) and *correlated stochastic noise* processes were considered since both have relevance to the neural recording problem. For example AWGN is frequently used in theoretical studies of neural recordings because *myelinated* axons are considered to be electrically isolated and therefore have no synaptic interactions within a nerve [21]. In theory, therefore, background noise in recordings made from hook or suction electrodes of large myelinated nerves could be approximated by a Gaussian noise process. In practice, however, *ephaptic* interactions between axons and crosstalk between amplifiers and recording equipment introduce various levels of correlation [13]. Furthermore any interference from other sources such as muscle activity will appear with some level of correlation. Therefore a correlated noise model was also considered and used to describe the background activity. It was generated using a dynamic *Ornstein-Uhlenbeck* (OU) process described by the difference equation:

$$OU_{t+dt} = OU_t - \frac{OU_t}{\tau} dt + d\Omega_t \quad (20)$$

Where OU_t is the value of the noise process at each time step, dt is the simulation time step, τ is the time constant of the process and Ω_t denotes a *Wiener* process [22]. A Weiner process is any process where $\Omega_0 = 0$, $E[\Omega_t] = 0$ and $\Omega_{t+dt} - \Omega_t \sim N(0, dt)$. Figure 7 illustrates the frequency content of the AWGN and OU noise processes from left to right respectively for purposes of comparison. The spectra were calculated using the FFT with time domain noise samples $100\,ms$ in length with a sample rate of $50\,kHz$ and $\tau = 0.01\,ms$. The spectrum of the OU noise process is inherently band limited and the noise power of the OU noise process falls below that of the AWGN model at a frequency of approximately $3\,kHz$.

C. *Results with AWGN and OU noise processes*

In the simulations described in this section both AWGN and OU noise processes were applied to APs *separately* using the following method:

- A single AP was simulated in each case with a diameter of $15\,\mu m$ at 21 degrees C, the sampling frequency was f_s = 500 kHz;
- The AP reference position was then calculated using (**a**) the point of maximum slope of the AP, (**b**) the maximum value, (**c**) the mid-point of the -3 dB points and (**d**) the centroid;
- In all simulations the SNR was swept from $+40\,dB$ to $-40\,dB$ in linear steps of $-1dB$;
- For each value of SNR the experiment was repeated 100,000 times and the pulse parameters (**a-d**) were computed each time and averaged;
- The mean and the standard deviation for the pulse parameters at each level of SNR were computed relative to a normalized reference position of 100 samples for the purposes of comparison and presentation.

The normalized plots shown in Figures 5 and 6 were obtained by applying these methods using AWGN. The mean locations and standard deviations of the pulse parameters are plotted for the four alignment methods (**a – d**). For the case of AWGN, in addition to applying the white noise directly, the noise was low-pass filtered representative of the first part of the spike sorting process. An 8th order Butterworth digital (IIR) filter with a cut-off frequency of 10 kHz was employed. Figures 5 and 6 are split into two pairs (5a, 5b & 6a, 6b) to accommodate this comparison.

The main feature of these results (i.e. for both mean and standard deviation) is that pulse location based on maximum slope (**a**) appears to be the most susceptible to

noise, degrading severely with an SNR less than about 30 dB. This is followed by the method based on the maximum value of the pulse (**b**), which fails with an SNR of approximately 0 dB. The methods based on the mean of the -3 dB points (**c**) and the centroid (**d**) perform much better, operating well with SNRs of about -10 dB. In these simulations the centroid method always outperformed the others.

In the second case white (i.e. uncorrelated) noise was replaced by correlated noise generated using the OU model defined above. The results are shown in Figures 8 and 9 for the mean and standard deviation respectively. The main effect of this change was that methods (**a**) and (**b**), which performed worst under AWGN (especially in the case where filtering was *not* employed), performed much better using OU noise. The overall effect was that the spread in performance of all four methods was much less than when AWGN was used, for both the mean and SD measures. This suggests that single point temporal measures such as methods (**a**) and (**b**) are particularly sensitive to high frequency noise components such as are present in AWGN. In addition both these methods rely on determining *turning points* in the time record, emphasising their vulnerability to noise in general and high frequency components in particular.

D. Power and area measurements

An analysis of the power and resource requirements for the proposed centroid filter was performed using dedicated hardware implementations constructed using an FPGA (Altera Cyclone II EP2C35F672C6N FPGA) and a CPLD (Altera MAX V 5M570ZF256C5N). These devices were chosen as rapid prototyping platforms both because of their low cost compared to a custom ASIC and because at lower sampling rates an FPGA or CPLD can consume less power than an equivalent processor-based implementation [23]. Furthermore the relatively low sampling rates associated with ENG, typically less than 100 kS/s, may allow a reduction in core operating voltages and thus a saving in static power consumption in an FPGA [24]. Additionally the use of a MAX V CPLD allowed for separate external regulation of the core and I/O voltage busses. This configuration permits more accurate power consumption measurements as the two domains can be easily separated. The designs were produced using the SystemVerilog hardware description language and RTL verification was performed using the QuestaSim environment (*QuestaSim 10.2c, Mentor Graphics Inc, Oregon, USA*).

The synthesis tools (*Quartus 13.0, Altera Corporation, Calif. USA*) reported an initial

maximum clock rate for the centroid filter of 25 MHz. The designed clock rate was chosen to be 500 kHz, representative of the fastest sampling rate used in recent acute experiments [25]. Power measurements were made using a current sense resistor (1 Ω) connected in-line with the DC power supply. Baseline power measurements were made using clock gating to isolate the filter structure, as well as to estimate the power consumed by the I/O and clock propagation circuitry. The CPLD implementation utilized separate core and I/O power supplies, both at 3.3 V, for more accurate measurements. The power requirements were measured at a clock frequency of 500 kHz. Resource measurements were taken from the synthesis tools after optimisation and fitting of the design. An overview of the resource and power requirements achieved for both methods, is provided in Table II. Power analysis tools (*Powerplay Power Analyser, Altera Corporation*) were used to predict the power consumption of the designs and the predicted values were in close agreement with measured data.

TABLE II

DEVICE UTILIZATION SUMMARY AND POWER CONSUMPTION f_{clk} = 500 kHz

Device	Parameter	Result
CPLD - 5M570ZF256C5N	LC Registers	330
	Logic Cells	442
	Current Consumption	0.09 mA
	Total Power (measured)	0.17 mW
	Total Power (estimated)	0.15mW
FPGA - EP2C35F672C6N	LC Registers	330
	Logic Cells	442
	Current Consumption	34.5 mA
	Total Power (measured)	62.2 mW
	Total Power (estimated)	61.5 mW

E. Verification

Test pattern waveforms were generated in MATLAB using the SFAP model Equation (1) before being sampled and transformed into a Qs0:7 fixed point format. The RTL codes used to produce designs for the two devices were identical. In addition to the sample memories a verification block was included within the design. This block contained the expected output from both filters and performed simple on-line comparisons on a sample-by-sample basis. The data files for both the sample memory and the verification block were transferred using a custom JTAG programming interface into the on-chip synchronous memories – implemented in 4K RAM blocks - of both the FPGA and the CPLD. A memory pointer was driven from a variable down

sampled clock (system clock speed was 500 kHz, down sampled from 10 MHz), this pointer was used to address the circular memory buffer, the returned samples being fed into the centroid filter. The clock control circuitry was used to activate individual modules to obtain accurate power measurements. The verification circuity confirmed correct operation of the centroid filter up to a maximum clock frequency of 24.5 MHz at which point the expected output deviated from the observed output, this frequency is in agreement with the predicted maximum clock frequency of 25 MHz.

5. DISCUSSION

A. General effects of noise on spike alignment

As outlined in the introduction, neural recording methods have developed greatly in recent years with the development of multi-electrode methodologies for a wide variety of applications. Of these methods, morphological spike sorting is amongst the most popular. Unfortunately one of the key stages of such spike sorting methods, the temporal alignment stage, is very susceptive to noise and interference [26]. This is because the methods employed to fix the position of individual spikes (APs) in time tend to depend on single point measurements of each spike and are therefore particularly sensitive to the effects of noise processes. This sensitivity was demonstrated in the simulations reported in Section 4 of this paper for both uncorrelated and correlated additive noise sources. The basic approach adopted in this paper is therefore to propose a method that does not rely on single point determinations of spike locations, but rather employing a measure that in some sense is an average of the whole AP. As we have demonstrated, this has the effect of reducing the sensitivity of the spike position measurement to additive noise, at least for uncorrelated sources.

B. Comparison of AWGN and OU

It is tempting to assume that because *myelinated* axons are generally considered to be electrically isolated and therefore have no synaptic interactions within a nerve that all noise sources can be considered to be *uncorrelated* (AWGN) [8], [21]. In practice, however, various interactions between axons and crosstalk between amplifiers and recording equipment introduce some correlation [13]. Therefore a correlated noise model was also considered in the simulation study, based on the Ornstein-Uhlenbeck process [22]. The effect of this change was dramatic on the two methods of spike

position determination based on single points in time. These were the point of maximum AP slope and the actual maximum point. This was not very surprising because in addition to being single point measures these methods require the calculation of a turning point in time, suggesting an enhanced sensitivity to high frequency noise components. This view was confirmed by lowpass filtering the signal after the application of AWGN. This is standard practice in spike sorting and the result was comparable to the effect of changing to a correlated source, which is naturally band limited (see Figures 5 & 6).

C. Benefits of the centroid filter

In all cases, and independent of the type of noise or interference applied to the recordings, the centroid proved to be a significantly more reliable metric for aligning the APs. Intuitively the centroid of the AP is a function of the energy (or area) contained within the AP itself, the addition of broadband AWGN with zero mean will have little if any effect on the amount of energy within the AP. Of course this simplification does not hold when considering finite time sequences or band-limited noise such as the OU process considered in this study.

In its most basic form computing the centroid introduces significant overhead, especially in comparison to the single sample measures. However it has been shown that the computation can be performed with minimal effort using a novel FIR filter structure. It has been shown that the application of a low pass filter to the raw signal improves the performance of the single sample methods for AP alignment and in this study an 8^{th} order Butterworth IIR lowpass filter with an upper cut-off frequency of 10 kHz was used.

When considering the implementation costs, the centroid filter, which uses only a single multiplier, will significantly out-perform a direct-form IIR filter of almost any order, as the number of multipliers required typically scales with the filter order. Table III details the relative implementation costs for both the centroid filter and the 8^{th} order Butterworth filter used in this study, the Butterworth filter consumes over five times the number of Logic Cells and nearly ten times as much power as the centroid filter. The centroid filter is an efficient and low power method for AP alignment that shows considerable resilience to noise even in SNRs less than – 10 dB. It should be noted that this method has applications in a wide range of areas wherever

phase information between different pulses is considered, for example in phase sensitive communication systems.

6. CONCLUSIONS

A new method has been described that significantly improves one of the key issues in neural spike sorting, i.e. the difficulty of correctly aligning action potentials (APs or 'spikes') in the time domain in the presence of noise. A novel real-time approach for spike alignment based on a *centroid* filter is described that provides an alternative to traditional spike alignment methods and substantially improves resilience to noise and resulting sampling jitter. We have validated the new methods by simulation using

TABLE III
DEVICE UTILIZATION SUMMARY AND POWER CONSUMPTION F_{clk} = 500 kHz

Method	Parameter	Result
Band-pass Filter (FPGA)	LC Registers	168
	Logic Cells	2533
	Current Consumption	343.7 mA
	Total Power (measured)	618 mW
	Total Power (estimated)	601mW
Centroid Filter (FPGA)	LC Registers	330
	Logic Cells	442
	Current Consumption	34.5 mA
	Total Power (measured)	62.2 mW
	Total Power (estimated)	61.5 mW

deterministic models of nerve signals (i.e. APs with added noise; both correlated and uncorrelated noise models are considered). In addition, we have shown that the new method lends itself particularly well to hardware realization and a power efficient solution is described that operates in real time on a single Altera Max V *complex programmable logic device* (CPLD). The technique has the potential to influence significantly the design of electrophysiological recording systems in the future.

7. References

[1] S. Gibson, J. W. Judy, and D. Markovi, "Spike Sorting, the first step in decoding the brain," *IEEE Signal Processing Magazine*, no. January, pp. 124–143, 2012.

[2] K. Sameshima and L. A. Baccala, "Trends in multichannel neural ensemble recording instrumentation," in *Methods for Neural Ensemble Recordings*, M. A. L. Nicholelis, Ed. CRC, 1999, pp. 47–60.

[3] K. H. Kim and S. J. Kim, "Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier.," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 10, pp. 1406–11, Oct. 2000.

[4] C. M. Stewart, S. D. Newlands, and A. A. Perachio, "Spike detection, characterization, and discrimination using feature analysis software written in LabVIEW.," *Comput. Methods Programs Biomed.*, vol. 76, no. 3, pp. 239–51, Dec. 2004.

[5] G. Zouridakis and D. Tam, "Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering," *Comput. Methods Programs Biomed.*, vol. 61, pp. 91–98, 2000.

[6] S. Gibson, J. W. Judy, and D. Markovic, "Spike Sorting: The First Step in Decoding the Brain: The first step in decoding the brain," *IEEE Signal Process. Mag.*, vol. 29, no. 1, pp. 124–143, Jan. 2012.

[7] J. Kaiser, "On a simple algorithm to calculate the energy of a signal," *Acoust. Speech, Signal Process. 1990. …*, vol. 2, no. 10, pp. 381–384, 1990.

[8] S. Mukhopadhyay and G. C. Ray, "A new interpretation of nonlinear energy operator and its efficacy in spike detection.," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 2, pp. 180–7, Feb. 1998.

[9] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials.," *Network*, vol. 9, no. 4, pp. R53–78, Nov. 1998.

[10] M. Abeles and M. G. Jr, "Multispike train analysis," *Proc. IEEE*, vol. 65, no. 5, 1977.

[11] P. Hartigan, "Algorithm as 217: Computation of the dip statistic to test for unimodality," *... R. Stat. Soc. Ser. C (Applied Stat.*, vol. 34, no. 3, pp. 320–325, 1985.

[12] J. Hartigan and P. Hartigan, "The dip test of unimodality," *Ann. Stat.*, 1985.

[13] D. A. Adamos, E. K. Kosmidis, and G. Theophilidis, "Performance evaluation of PCA-based spike sorting algorithms.," *Comput. Methods Programs Biomed.*, vol. 91, no. 3, pp. 232–44, Sep. 2008.

[14] C. Zhang, X. Zhang, M. Q. Zhang, and Y. Li, "Neighbor number, valley seeking and clustering," *Pattern Recognit. Lett.*, vol. 28, no. 2, pp. 173–180, Jan. 2007.

[15] N. Donaldson, R. Rieger, M. Schuettler, and J. Taylor, "Noise and selectivity of velocity-selective multi-electrode nerve cuffs.," *Med. Biol. Eng. Comput.*, vol. 46, no. 10, pp. 1005–18, Oct. 2008.

[16] S. Narasimhan, X. Wang, and S. Bhunia, “Implantable electronics: emerging design issues and an ultra light-weight security solution.,” *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2010, pp. 6425–8, Jan. 2010.

[17] B. W. Metcalfe, D. J. Chew, C. T. Clarke, N. N. Donaldson, and J. T. Taylor, “An Enhancement to Velocity Selective Discrimination of Neural Recordings: Extraction of Neuronal Firing Rates,” in *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, vol. 2014, pp. 4111–4114.

[18] B. W. Metcalfe, D. J. Chew, C. T. Clarke, N. de N. Donaldson, and J. T. Taylor, “A new method for spike extraction using velocity selective recording demonstrated with physiological ENG in Rat.,” *J. Neurosci. Methods*, vol. 251, pp. 47–55, Aug. 2015.

[19] M. . Protter and C. . Morrey, “College calculus with analytic geometry,” 1977, p. 526.

[20] N. Dalkiliç and F. Pehlivan, “Comparison of fiber diameter distributions deduced by modeling compound action potentials recorded by extracellular and suction techniques.,” *Int. J. Neurosci.*, vol. 112, no. 8, pp. 913–30, Aug. 2002.

[21] J. Segundo, J. Vibert, K. Pakdaman, M. Stiber, and O. D. Martinez, “Noise and the neurosciences: a long history, a recent revival and some theory,” *Orig. Brain Self Organ.*, p. 299, 1994.

[22] L. M. Ricciardi and L. Sacerdote, “The Ornstein-Uhlenbeck process as a model for neuronal activity,” *Biol. Cybern.*, vol. 35, no. 1, pp. 1–9, Mar. 1979.

[23] P. Kwan and C. Clarke, “FPGAs for improved energy efficiency in processor based systems,” *Adv. Comput. Syst. Archit.*, pp. 440–449, 2005.

[24] C. T. Chow, L. S. M. Tsui, P. H. W. Leong, W. Luk, and S. J. E. Wilton, “Dynamic voltage scaling for commercial FPGAs,” *Proceedings. 2005 IEEE Int. Conf. Field-Programmable Technol. 2005.*, pp. 173–180, 2005.

[25] B. Metcalfe, D. Chew, C. Clarke, N. Donaldson, and J. Taylor, “Fibre-selective discrimination of physiological ENG using velocity selective recording: Report on pilot rat experiments,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 2645–2648.

[26] H. K. Jung, J. H. Choi, and T. Kim, “Solving alignment problems in neural spike sorting using frequency domain PCA,” *Neurocomputing*, vol. 69, no. 7–9, pp. 975–978, Mar. 2006.

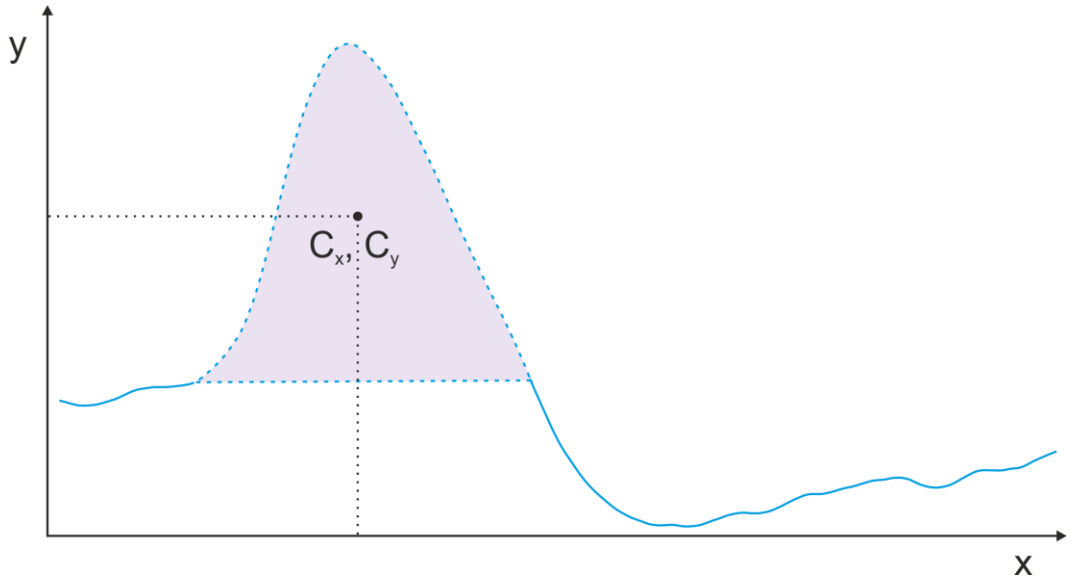


Figure 1: The concept of transforming a waveform into a polygon. The waveform (solid line) has been partitioned into a shaded area that may be treated as a plane polygon with centroid located at C_x, C_y.

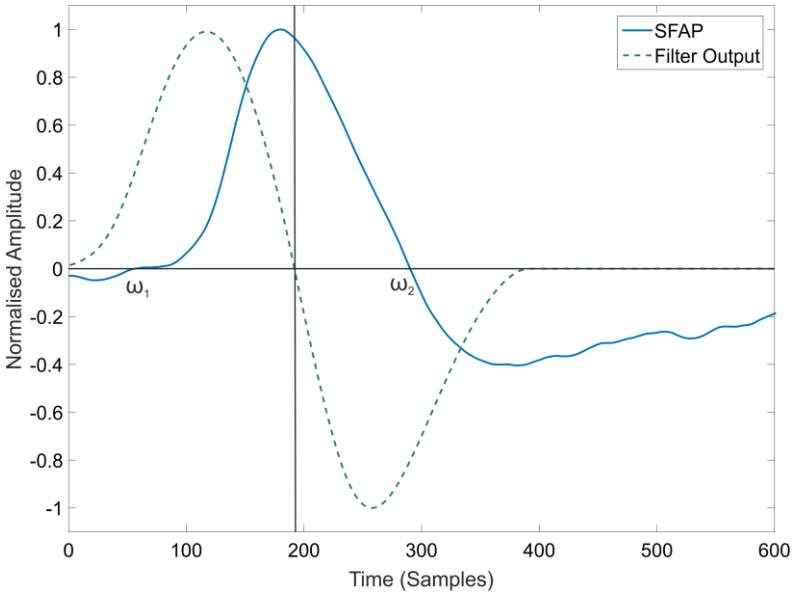


Figure 2: The application of the centroid filter to a single channel of data containing a realistic AP. In this case the width of the centroid filter was chosen to be N = 100 samples, or approximately the width of the positive phase of the AP. The solid line represents a discrete time version of the AP, and the dashed line the filter output $y[n]$. The vertical marker is set at the negative-going zero crossing of the filter output and is located at the centroid of the TM function.

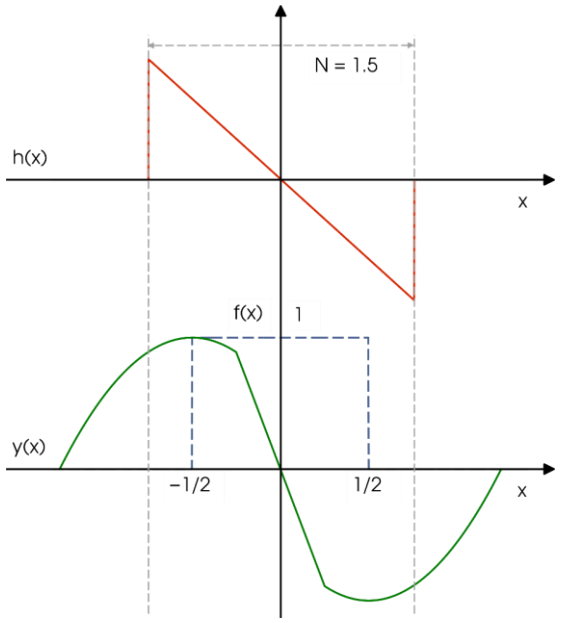


Figure 3: An example to illustrate the calculation of the centroid of an AP using the FIR filter *h(x)*. For clarity, *h* is represented by the continuous-time function shown in the upper plot in the Figure while the 'top hat' function shown in the lower plot represents the AP. Application of the convolution and shift functions results in the output function *y(x)* whose zero crossing corresponds with the centroid of the 'top hat' function. Note *y(x)* is shown in normalised form here for simplicity.

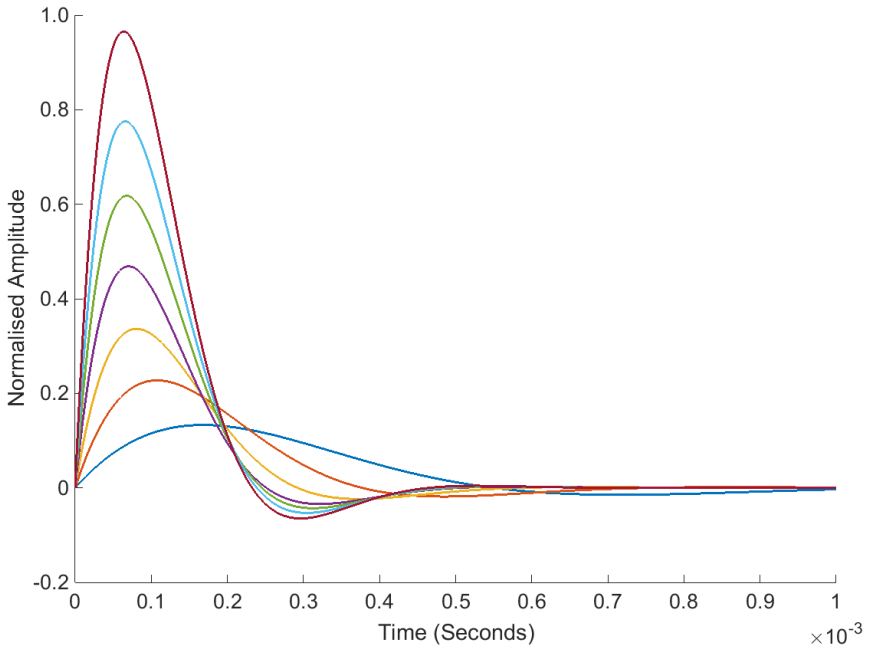
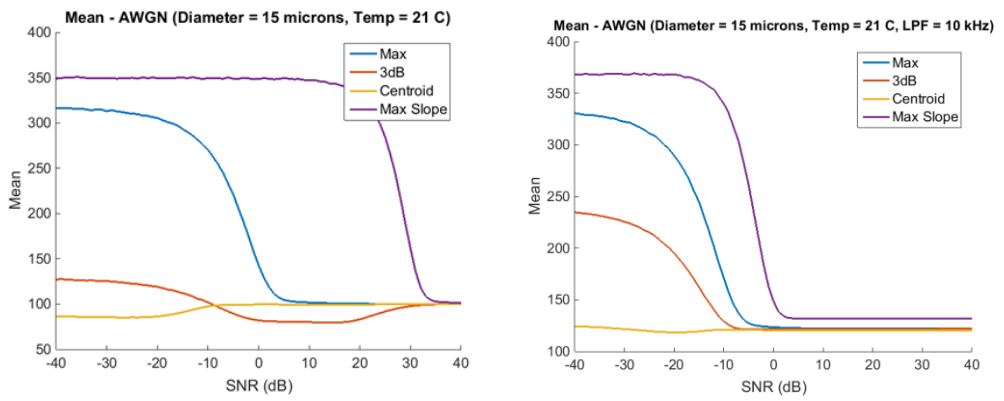
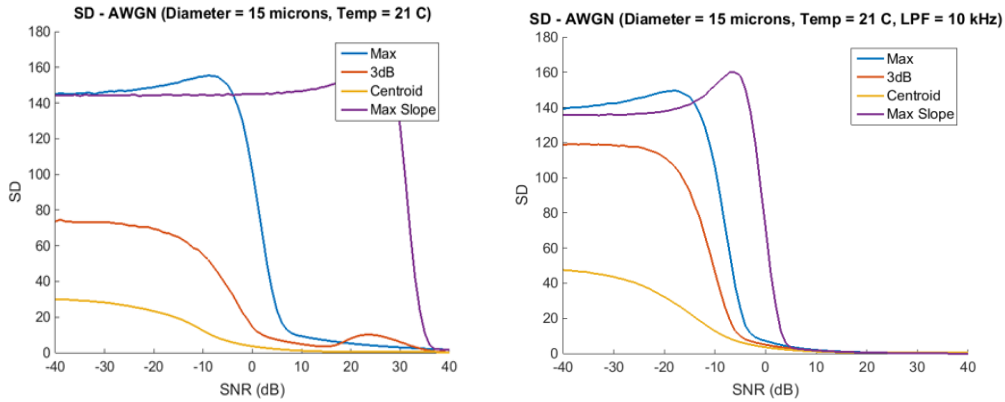


Figure 4: Seven different SFAPs generated using the damped sinusoid model given in Equation (1). The diameters are 5 (blue), 7 (red), 9 (yellow), 11 (mauve), 13 (green), 15 (blue) & 19 microns corresponding to conduction velocities of 9.3, 13, 16.7, 20.4, 24.2, 28, 31.6 & 35.3 m/s respectively.



(a) (b)

Figure 5: The mean locations for the pulse parameters in the presence of additive white Gaussian noise with SNR from 40 dB to -40 dB. In Fig 5(a) no filtering is applied to the noise whereas in (b) a lowpass filter with a cut-off frequency of 10 kHz was used which is representative of the first stage of the spike sorting process. The simulated SFAP is representative of an axon with diameter 15 μm at a temperature of 21 Celsius. The plots have been normalised to a theoretical mean of 100.



(a) (b)

Figure 6: The standard deviations of the pulse parameters in the presence of additive white Gaussian noise with SNR from 40 dB to -40 dB. In Figure 6(a) no filtering is applied to the noise whereas in (b) a lowpass filter with a cut-off frequency of 10 kHz was used which is representative of the first stage of the spike sorting process. The simulated SFAP is representative of an axon with diameter 15 μm at a temperature of 21 Celsius. The plots have been normalised to a theoretical value of zero.

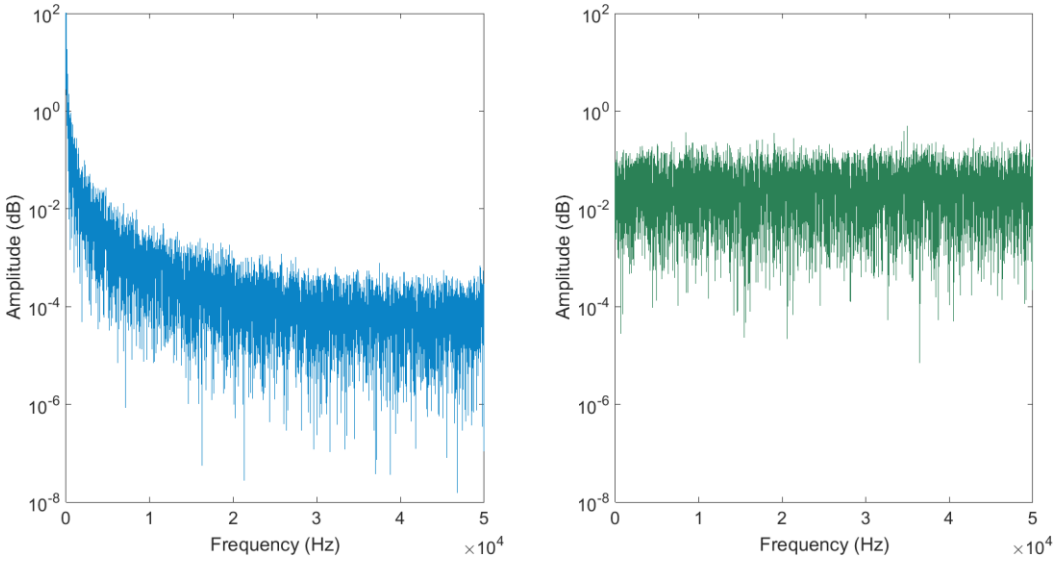


Figure 7: Frequency spectra for AWGN (**left inset**) and OU (**right inset**) noise processes computed using the FFT over a window of 100 ms. The time constant for the OU process (τ) was 0.01.

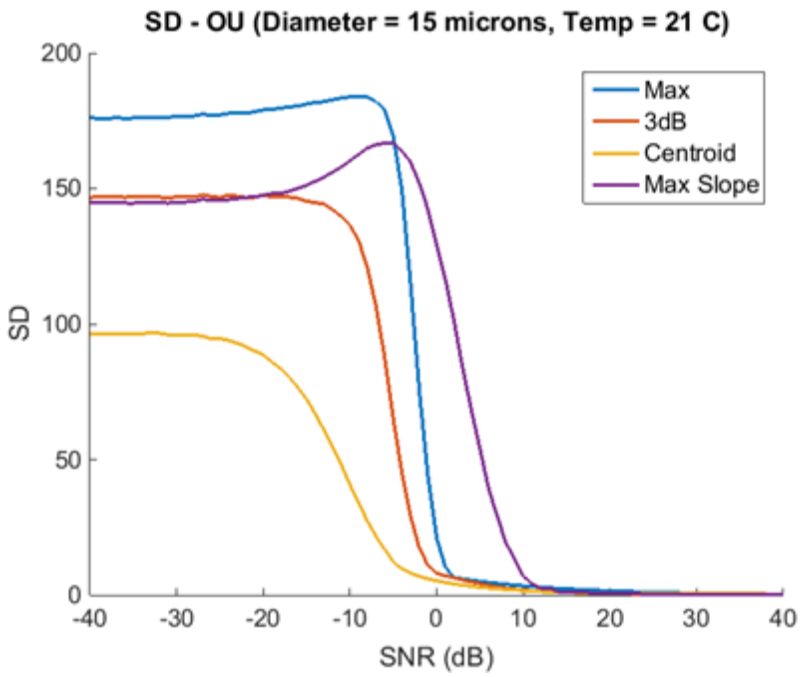


Figure 9: The standard deviations of the pulse parameters in the presence of Orstein-Uhlenbeck noise with SNR from 40 dB to -40 dB. The simulated SFAP is representative of an axon with diameter 15 μm at a temperature of 21 Celsius. The plots have been normalised to a theoretical value of zero.

plots have been normalised to a theoretical mean of 100.